University Admit Eligibility Predictor

.

Team ID : PNT2022TMID34617

Team Size : 4

Team Leader : BERLIN JINO D
(962219104038)

Team member : ANAND RAJ V

(962219104015)

Team member : DINESH RAM G S
(962219104049)

Team member : FLOREN DARIOUS K

(962219104054)

Abstract:

This is the literature review of University Student admission predictor application in the field of Data Science .The Objective for developing this application is to help students to fill their preferences at the time of option-entry process accurately and to ease of making better choices of college before allotment. This review examines various previous works done on Student admission predictor by using Data mining and ML technologies. The aim of this research is to develop a system using machine learning algorithms . It will help the students to identify the chances of their application to an university being accepted . Also it will help them in identifying the universities which are best suitable for their profile and also provide them with the details of those universities .A simple user interface will be developed for the users to access the SAP system.

INTRODUCTION:

University educations are an essential part of most people's preparation for working life. An admission to university is therefore became an important challenging topic. Hence, effective university admission prediction services are needed for helping students to enter the right university college. However, due to the huge numbers of students required to attend the university every year, this decision making process became a very complex problem. Since, this process is not merely relying on student test scores but also depends on students' backgrounds and other qualifications weighting criteria that correlate to the performance of their tertiary education. This review involves the related works done on Student admission predictor to Universities and colleges based on Machine learning, Data Mining and Analysis methodologies.

Data Science: History and Present:

As a method of prediction previously many years statistics was used. Until recent years these statistics have been transformed into data analysis. Ten to twenty years ago this data analysis was further urged to develop into a new stream known as data science .As maximum amount of data was stored electronically extracting and analyzing it itself became a science. Around fifty years ago Jhon Tucky evolved statistics into the term data analysis (Donoho, 2015). Later around twenty years ago this data analysis was further evolved into new stream known as data science by John Chambers, Bill Cleveland and Leo Breiman (Donoho, 2015). Data Mining has been the oldest form of technique for data science. Various algorithms are being developed in order to aid this process. The evolution of these algorithms date back to 1957. In 2006 IEEE announced top 10 algorithms used for data data mining (Wu et al., 2007):

University Admit Eligibility Predictor:

Review of Related works:

Bayesian Networks were used by Thi et al. (2007) to create a decision support system for evaluating the application submitted by international students in the university. This model was designed to predict the performance of the aspiring students by compar ing them with the performance of students currentl studying in the university and had similar profile during their application. In this way based on the current students profile the model predicted whether the aspiring student should be granted admission to the university.

Abdul Fatah S; M (2012) developed a model that can provide the list of universities/colleges where the which best suitable for a student based on their academic recordsand college admission criteria. The model was developed by

applying data mining tech niques and knowledge discovery rules to the already existing in-house admission prediction system of the university.

Data mining methods used here are

a)Web Mining:

Web usage mining performs mining on student's web data, particularly data stored in logs managed by the web servers.

b)Recommendation Using KDAR:

One of the best-known examples of DM in recommender systems is the discovery of association rules, or item-to-item correlations.

c)Students Data Clustering:

Clustering techniques work by identifying groups of students who appear to have similar preferences.

d)Students Classification:

Classifiers are general computational models for assigning a category to an input.

Mane (2016) conducted a similar research that predicted the chance of a student getting admission in college based on their Senior Secondary School,

Higher Secondary School and Common Entrance Examination scores using the pattern growth approach to association rule mining. The performance of both the models was good the only drawback was the problem statement was single university-centric.

Mishra and Sahoo (2016) conducted a research from a university point of view to predict the likelihood of a student enrolling in the university after the have enquired about of courses in the university. They used K-Means algorithm for

clustering the stu dents based on different factors like feedback, family income, family occupation, parents qualification, motivation etc. to predict if the student will enroll at the university or not. Depending upon the similarity of the attributes among the students they were grouped into clusters and decisions were made. The objective of the model was to increase the enrolment of the students in the university.

Eberle et al. (n.d.) used machine learning and predictive modelling to develop a model that to evaluate the admission policies and standards in the Tennessee Tech Uni versity. A well know version of the C4.5 algorithm, J48 was used to create the model. Like the models mentioned above they used the different factors of the student profile to evaluate the chances of their admission in the university. The model worked well in predicting the true positive scenarios where the student was had good profile to secure the admission, but it failed in efficiently identifying the true negatives because of which student who does not satisfy the defined criteria.

GRADE system was developed by Waters and Miikkulainen (2013) to support the admission process for the graduate students in the University of Texas Austin Department of Computer Science. The main objective of the project was to develop a system that can help the admission committee of the university to take better and faster decisions. Lo gistic regression and SVM were used to create the model, both models performed equally well and the final system was developed using Logistic regression due to its simplicity. The time required by the admission committee to review the applications was reduced by 74% but human intervention was required to make the final decision on status if the

application. Nandesh waretal. (2014) created a similar model to predict the enrolment of the student in the university based on the factors like SAT score, GPA score, residency race etc. The Model was created using the Multiple Logistic regression algorithm, it was able to achieve accuracy rate of 67%.

Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S, Proposed a University Admission Predictor (UAP) system. They have tried to determine the predictions using various machine learning algorithms such as KNN, Decision Trees, Ridge Regression etc to predict the admission chances using the variables such as GRE, TOEFL, GPA, SOP, LOR etc. During the research they have tried to compare the algorithms mentioned above for choosing the best for their proposed UAP system. The proposed system is able to predict the admission chances with 79% of average accuracy.

Md. Imdadul Hoque, Abul kalam Azad, Mohammad Abu Hurayra Tuhin, Zayed Us Salehin,proposed a University Student Result Analysis and Prediction system by implementing the Decision Tree machine learning algorithm. The system can be used to predict the student results by analyzing their
previous year/semester marks and the model also outlines the subjects in which the student is lacking. The system tries to help students find out the area or
the subject where a student needs to put more focus so that he/she can avoid poor achievement and perform better. The data used for the prediction model is taken from the university database as well as through surveys done using google form. The data set contains information such as gender, extracurricular activities, number of tuition, programming skills, class test marks, assignment marks, attendance and previous GPA. The study tried to relate the performance prediction closely with CGPA. To train the data model weka tools was used. The

objective of the study also revolves around the comparison of three different types of decision trees as well and concludes that the J48 decision tree algorithm outperforms the other two algorithms achieving the highest accuracy.

Machine Learning technique used:

Several machine learning algorithms have been used by several authors in the above mentioned literature review. In this section we will be seeing those methodologies and how these classifiers works. This section also deals with the inner workings of the listed methodologies.

Linear Regression:

Linear regression is a regression model which uses supervised learning approach of machine learning. As the name suggests, a regression task is performed with this algorithm and a model is created based on the regression to derive meaningful contexts from a given set of data. It is one of the basic classifier used in machine learning. The algorithm is used to find out the relationship between dependent and independent variables.

Logistic Regression:

Logistic regression is similar to linear regression but instead of a linear response given by the linear regression, the logistic regression has a binomial response variable. In the logistic regression we can have more than 2 continuous explanatory variables and it's easier to handle those variables simultaneously. In the case of more than one explanatory variable, logistic regression is used to calculate the odds ratio. With the fact that the outcome variable is binomial, the technique is very similar to multiple linear regression.

Naive Bayes Classifier:

The Naive Bayes classifier consist of a wide collection of algorithms which are based on Bayes' Theorem. Thus it's not a single algorithm but a family of algorithms in which all of them share a common principle, i.e every pair of a selected feature or characteristic is classified as independent of each other.

Random Forest:

A Random Forest is an ensemble learning method that can execute both regression and classification tasks by combining several decision trees and a technique known as Bootstrap and Aggregation, also known as bagging. The basic principle is to use several decision trees to determine the final production rather than relying on individual decision trees. Random Forest's foundation learning structures are multiple decision trees. We randomly select rows and features from the dataset to create sample datasets for each model.

Decision Tree:

The decision tree is the most effective and widely used classification and prediction method. A Decision tree is a tree structure that looks like a flowchart, with each internal node representing a test on an attribute, each branch representing the test's result, and each leaf node (terminal node) holding a class name.

CONCLUSION:

The numbers of the students seeking further higher education abroad keeps increasing every year. And this number won't go down in upcoming years either. In our technological and competitive world the students constantly feel the need to upskill themselves and fare better than their peers and competitors everyday and thus after their under graduation, a vast number of students tend to leave for higher education abroad. To facilitate an easy migration, and to provide them with a tool which can help them shortlist their potential college/university based on various parameters will not just empower them in finding their preferred institutions but also ensure that it saves their precious time as well as their hard earned money at the same time. Many researchers have tried to come up with ideas using machine learning, data mining, and generic algorithm from time to time, each with their own benefits and drawbacks. If we go through the research done till date, the success rate of hybrid systems with the implementations of more than one algorithm concurrently is higher with low error rates, as the work done by each algorithm in any hybrid system relies on each of them only for the specific task and hence later collaboration ensures higher accuracy. In this regard, the stacked ensemble approach, which itself is based on an ensemble approach which tries to bring different algorithms together to gain better accuracy, fares better than all the other approaches discussed and reviewed in this paper. An improvement can also be made and the system can always be made more efficient and accurate using NLP or natural language processing methods which can help rate the quality of the written essays and paragraphs during the exams, so that those features can also help narrow down the chances of admissions over time.

REFERENCES:

1. Chithra Apoorva D A, Malepati ChanduNath, Peta Rohith, Bindu Shree.S, Swaroop.S, "Prediction for University Admission using Machine

Learning" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-6 March 2020

2. Md. Imdadul Hoque, Abul kalam Azad, Mohammad Abu Hurayra Tuhin, Zayed Us Salehin,"University Students Result Analysis and Prediction System by Decision Tree Algorithm" Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 3, 115-122 (2020) ,DOI: 10.25046/aj050315

3. Mane, R. V. (2016). Predicting Student Admission decisions by Association Rule Mining with Pattern Growth Approach, pp. 202–207

4. Eberle, W., Simpson, E., Talbert, D., Roberts, L. and Pope, A. (n.d.). Using Mach

5. Abdul Fatah S; M, A. H. (2012). Hybrid Recommender System for Predicting College Admission, pp. 107–113.